

图像处理方法和图像样本分类方法、装置和存储介质

本申请要求于 2024 年 1 月 04 日提交中国专利局、申请号为 202410010706.X、发明名称为“图像处理方法和图像样本分类方法、装置和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请涉及图像处理技术领域，更具体地涉及一种图像处理方法、一种图像样本分类方法、一种图像处理装置、一种图像样本分类装置、一种电子设备及一种存储介质。

背景技术

随着机器视觉技术的发展，在很多应用场景中需要使用图像来实现各种生产目的。

10 例如，在工业领域中，特别是晶圆生产等高精尖的子领域，产品的良率要求极高。因此，期待利用产品图像来实现产品的缺陷检测。但是当产品产量不足或没有大批量生产时，缺陷产品数量有限。由此，缺陷产品图像的数量也有限，缺陷产品图像样本集的收集难度极大。由于图像样本集中的图像数量有限，其通常难以完美表示缺陷产品的特征，这导致了在缺陷自动化检测时，检测准确率非常低。

为此，迫切需要一种图像处理方法来处理有限的图像，使其更理想地代表其中目标对象的特征。

15 发明内容

为了至少部分地解决现有技术中存在的问题，根据本申请的第一个方面，提供了一种图像处理方法，包括：利用特征提取网络对多个第一图像分别进行特征提取，以得到多个第一特征向量；对多个第一特征向量进行聚类处理，以获取代表多个第一图像的、第一预设数量的第一向量组和代表每个第一向量组的第一代表特征向量；对于每个第一向量组，计算该第一向量组的分布数据，其中分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差；基于参考图像的、第一预设数量的参考向量组各自的分布数据，校正多个第一向量组的分布数据，其中，参考向量组与第一向量组一一对应；以及将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，作为多个第一图像的特征表示数据。

在一种可能的实现方式中，特征提取网络包括级联的两个特征提取网络。

25 在一种可能的实现方式中，分布数据为权重，计算该第一向量组的分布数据，包括：计算该第一向量组中的所有第一特征向量与该第一向量组中的第一代表特征向量之间的距离；计算第一数量与多个第一特征向量的总数量之间的比值，以作为该第一向量组的权重；其中，第一数量是与该第一向量组中的第一代表特征向量之间的距离小于距离阈值的第一特征向量的数量。

在一种可能的实现方式中，分布数据包括权重，基于参考图像的、第一预设数量的参考向量组各自

的分布数据，校正多个第一向量组的分布数据，包括：从第一预设数量的第一向量组的权重中选取第二预设数量的权重，其中所选取的权重大于未被选取的权重；利用第一参考向量组的权重，对应地替换所选取的权重。

在一种可能的实现方式中，处理方法还包括：计算多个第一特征向量的统计值；利用特征提取网络
5 对多组参考图像中每个参考图像进行特征提取，以得到多个参考特征向量，其中，不同组的参考图像属于不同类型；针对每组参考图像，对该组参考图像的参考特征向量进行聚类处理，以获取代表该组参考图像的、第一预设数量的第二向量组和代表每个第二向量组的第二代表特征向量；对于每个第二向量组，计算该第二向量组的分布数据；针对每组参考图像，计算该组参考图像的参考特征向量的统计值；根据多个第一特征向量的统计值以及每组参考图像的参考特征向量的统计值，计算多个第一图像与每组参考
10 图像的距离；根据多个第一图像与每组参考图像的距离，从多组参考图像中选择一组参考图像，并将代表所选择的该组参考图像的第二向量组确定为参考向量组。

根据本申请另一方面，还提供一种图像样本分类方法，包括：

利用特征提取网络对图像样本进行特征提取，以得到样本特征向量；

基于样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，其中，
15 每组第一图像的特征表示数据是利用上述图像处理方法获取的，不同组的第一图像属于不同类型。

在一种可能的实现方式中，基于样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，包括：

根据样本特征向量与每组第一图像的特征表示数据，计算每组第一图像与图像样本之间的第一距离，并得到多个第一距离；

20 将多个第一距离中的最小值所对应的一组第一图像所属的类型确定为图像样本所属类型。

在一种可能的实现方式中，第一图像是产品的图像，第一图像所属类型根据产品的缺陷的类型来确定。

根据本申请另一方面，还提供一种图像处理装置，包括

第一特征提取模块，用于利用特征提取网络对多个第一图像分别进行特征提取，以得到多个第一特
25 征向量；

聚类模块，用于对多个第一特征向量进行聚类处理，以获取代表多个第一图像的、第一预设数量的第一向量组和代表每个第一向量组的第一代表特征向量；

第一计算模块，用于对于每个第一向量组，计算该第一向量组的分布数据，其中分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差；

校正模块，用于基于参考图像的、第一预设数量的参考向量组各自的分布数据，校正多个第一向量组的分布数据，其中，参考向量组与第一向量组一一对应；

第二计算模块，用于将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，作为多个第一图像的特征表示数据。

5 根据本申请另一方面，还提供一种图像样本分类装置，包括：

第二特征提取模块，用于利用特征提取网络对图像样本进行特征提取，以得到样本特征向量；

分类模块，用于基于样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，其中，每组第一图像的特征表示数据是利用上述图像处理方法获取的，不同组的第一图像属于不同类型。

10 根据本申请另一方面，还提供一种电子设备，包括处理器和存储器，存储器中存储有计算机程序指令，计算机程序指令被处理器运行时用于执行上述图像处理方法和/或上述图像样本分类方法。

根据本申请另一方面，还提供一种存储介质，在存储介质上存储了程序指令，其程序指令在运行时用于执行上述图像处理方法和/或上述图像样本分类方法。

根据上述技术方案，利用参考图像的多个参考向量组各自的分布数据，校正第一图像的第一向量组的分布数据，并根据校正后的分布数据和每个第一向量组的第一代表特征向量作为多个第一图像的特征表示数据。由于校正后的分布数据和每个第一向量组的第一代表特征向量能够准确代表多个第一图像中目标对象的特征。从而，利用这些特征表示数据可以快速且准确地实现对同类型的图像的自动分类。并且，上述方案在第一图像有限的场景下也依然可以适用，方案的适用范围更广泛。

在发明内容中引入了一系列简化形式的概念，这将在具体实施方式部分中进一步详细说明。本申请内容部分并不意味着要试图限定出所要求保护的技术方案的关键特征和必要技术特征，更不意味着试图确定所要求保护的技术方案的保护范围。

以下结合附图，详细说明本申请的优点和特征。

附图说明

本申请的下列附图在此作为本申请的一部分用于理解本申请。附图中示出了本申请的实施方式及其描述，用来解释本申请的原理。

图 1 示出根据本申请一个实施例的图像处理方法的示意性流程图；

图 2 示出根据本申请一个实施例的特征提取网络的示意性框图；

图 3 示出根据本申请一个实施例的图像样本分类方法的示意性流程图；

图 4 示出根据本申请另一实施例的图像样本分类方法的流程图；

图 5 示出根据本申请实施例的图像处理装置的示意性框图；

图 6 示出根据本申请实施例的图像样本分类装置的示意性框图；

图 7 示出根据本申请实施例的电子设备的示意性框图。

具体实施方式

5 在下文的描述中，提供了大量的细节以便能够彻底地理解本申请。然而，本领域技术人员可以了解，如下描述仅示例性地示出了本申请的优选实施例，本申请可以无需一个或多个这样的细节而得以实施。此外，为了避免与本申请发生混淆，对于本领域公知的一些技术特征未进行详细描述。

 为了至少部分地解决上述技术问题，根据本申请的一个方面，提供了一种图像处理方法。图 1 示出根据本申请一个实施例的图像处理方法 100 的示意性流程图。如图所示，该方法 100 包括步骤 S120、步
10 骤 S130、步骤 S140、步骤 S150 和步骤 S160。

 步骤 S110，利用特征提取网络对多个第一图像分别进行特征提取，以得到多个第一特征向量。

 第一图像可以是任意合适的对象的图像，本申请不对其进行限制。示例性地，第一图像是产品的图
像，第一图像所属类型根据产品的缺陷的类型来确定。将本申请的方法应用于对产品的检测，可以在新
产品缺陷样本难以收集的情况下，也可以实现对产品的缺陷的精准分类，并且大幅降低人工成本，提高
15 生产效率，且检测准确度进一步提高。在未示出的实施例中，本申请的图像样本分类方法也可以应用于
其他场景，例如人脸识别等。因此，第一图像也可以是其他合适的图像，第一图像所属类型也可以是例
如人物、景观等各种对象的合适的分类类型。为便于描述，下文继续以第一图像是产品的图像，第一图
像所属类型是产品的缺陷类型为例进行展开阐述。

 以对象是晶圆为例。多个第一图像可以是晶圆新产品的缺陷样本。这些第一图像可以是经人工标注
20 缺陷类别的缺陷图像。第一图像的数目可以是任意的，本申请不对其进行限制。可以理解，晶圆新产品
的缺陷样本可能较少。多个第一图像可以是被标注为一种缺陷类型的图像。例如，每种缺陷类型选取 50
个样本，则第一图像的数目即为 50。步骤 S110 至步骤 S150 可以是针对每一种缺陷类型的 50 个第一图
像执行的。

 第一图像既可以是灰度图像，也可以是彩色图像。其可以是任意合适尺寸、合适分辨率的图像，也
25 可以是满足预设要求的图像。可以采用任何现有的或者未来的图像采集方式获取第一图像。第一图像可
以是图像采集装置直接采集到的原始图像，也可以是对原始图像进行预处理操作后的图像。该预处理操
作可以包括为了改善图像的视觉效果，提高图像的清晰度，或是突出图像中的某些特征等便于特征提取
的所有操作。例如，预处理操作可以包括滤波等去噪操作，也可以包括诸如对图像整体的灰度、对比度、
亮度等不影响特征提取的图像调整操作。

根据本申请实施例的特征提取网络可以是任意合适的图像特征提取网络，本申请不对其进行限制，包括但不限于卷积神经网络、残差网络、密集连接网络等。示例性而非限制性地，特征提取网络包括级联的两个特征提取网络。例如，特征提取网络可以是级联的两个卷积神经网络或级联的两个循环神经网络。

5 图 2 示出根据本申请一个实施例的特征提取网络的示意性框图。如图 2 所示，特征提取网络可以是经训练的特征提取与分类网络中的部分网络。特征提取网络可以包括两个级联设置的 **resnet18** 特征提取网络。特征提取与分类网络由该两个级联的 **resnet18** 特征提取网络以及一个分类网络组成。以上述对象是晶圆，第一图像是新产品的小批量缺陷样本为例。特征提取与分类网络可以是利用大批量的晶圆产品的缺陷样本训练过的。例如，可以使用以前生产的或者当前正在生产的、与上述晶圆新产品相似度较高的
10 的大批量产品的缺陷数据，可以通过人工对缺陷数据进行分类，并标注类别标签，作为特征提取与分类网络的训练样本。利用这些大批量的训练样本对特征提取与分类网络的整体进行训练。可以理解，利用这些训练样本对特征提取分类网络整体进行训练的同时，特征提取网络也得到了训练。从而可以使得特征提取网络针对当前的晶圆产品的图像具有较好的缺陷特征提取能力。示例性地，可以预先利用 **ImageNet** 的图像数据集单独训练 **resnet18_1** 网络。之后才利用上述大批量的训练样本对特征提取与分类
15 网络进行整体训练，从而可以提高训练效率和训练效果。

可以理解，利用上述经训练的级联的两个特征提取网络对多个第一图像进行特征提取，可以提高特征的表达能力和判别能力，从而可以提高特征提取的精度。

示例性地，第一图像的尺寸例如是 **512*512**，则第一特征向量可以是 **512*1** 的特征向量。以每种缺陷类型选取 50 个样本为例，针对每一种缺陷类型的 50 个第一图像进行特征提取，则可以得到 50 个特征向量，并可以将这 50 个特征向量作为表示当前类型缺陷的多个第一特征向量。假设总体的缺陷样本有 10 种缺陷类型的话，那就可获得 10 个第一特征向量集。每个第一特征向量集对应一个缺陷类别。每个第一特征向量集中包括多个 **512*1** 的第一特征向量。
20

步骤 **S120**，对多个第一特征向量进行聚类处理，以获取代表多个第一图像的、第一预设数量的第一向量组和代表每个第一向量组的第一代表特征向量。

25 可以理解，此步骤可以针对每个第一特征向量集中的多个第一特征向量执行。聚类处理是一种具有代表性的无监督分类，可以对每个第一特征向量集中的第一特征向量进行分类。可以采用任意合适的聚类算法实现此步骤。包括但不限于 **K** 均值聚类 (**K-means**)、层次聚类 (**Hierarchical Clustering**) 等。

以 **K-means** 算法为例。具体地，首先可以选择合适的聚类中心（代表性的数据点）作为初始聚类中心。然后可以计算第一特征向量之间的距离：根据选定的距离度量（如欧氏距离或余弦相似度等），计

算每个第一特征向量与其他第一特征向量之间的距离或相似性。之后，可以根据计算得到的距离或相似性，将每个第一特征向量分配到与其最近的聚类中心所对应的类别中。接着，对于每个类别，可以重新计算其第一特征向量的中心或代表性点，可以是第一特征向量点的平均值、中位数等。迭代优化，使得聚类中心更能代表所属类别的第一特征向量。重复进行分配第一特征向量和更新聚类中心的过程，直到

5 达到终止条件，如固定迭代次数、聚类中心变化小于某个阈值等。得到最终的聚类结果，即每个第一特征向量所属的类别。由此，可以将同一类别的第一特征向量作为一个第一向量组。每个第一向量组组间差异较大，而组内差异较小。

第一预设数量可以是任意合适的值，其可以根据样本总量和样本分布情况进行任意设置。示例性地，第一预设数量可以是 5、4、3 等。以第一预设数量是 5 为例。假设每个第一特征向量集中均包括 50 个

10 第一特征向量。在此步骤中，可以利用 K 均值聚类算法将每个第一特征向量集中的 50 个第一特征向量分为 5 个组，得到 5 个第一向量组。在聚类处理完成后，可以直接将每一第一向量组中的聚类中心作为该第一向量组的第一代表特征向量，也可以计算每个第一向量组中所包含的各个第一特征向量的均值向量作为该第一向量组的第一代表特征向量。由此，针对每个第一特征向量集可以划分为 5 个第一向量组，并可以得到与 5 个第一向量组一一对应的 5 个第一代表特征向量。

15 步骤 S130，对于每个第一向量组，计算该第一向量组的分布数据。其中分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差。

示例性地，分布数据包括权重。在上述示例中，对于对应每个缺陷类别的 50 个第一图像的第一特征向量集，可以通过聚类得到 5 个第一向量组。然后，可以采用多种合适的方法计算第一向量组的权重。

在一个示例中，可以直接将每个第一向量组中的第一特征向量的数量与全部第一特征向量的总数量的比值作为第一向量组的权重。例如，对于每个缺陷类别的 50 个第一图像的第一特征向量集，如果聚类出的 5 个第一向量组中的第一特征向量的数目分别 15、10、10、8、7，则可以确定这 5 个第一向量组的权重分别是 0.3、0.2、0.2、0.16、0.14。在另一个示例中，计算该第一向量组的分布数据，包括：计算该第一向量组中的所有第一特征向量与该第一向量组中的第一代表特征向量之间的距离；计算第一数量与多个第一特征向量的总数量之间的比值，以作为该第一向量组的权重。其中，第一数量是与该第一

20 向量组中的第一代表特征向量之间的距离小于距离阈值的第一特征向量的数量。例如，对于一个第一向量组，其中包括 20 个第一特征向量，则可以确定该第一向量组的权重是 0.4。

换言之，可以设置一个距离阈值，对每个第一向量组进行判断：将计算得到的该第一向量组中的所有第一特征向量与该第一向量组中的第一代表特征向量之间的距离与距离阈值进行比较，得到距离小于距离阈值的第一特征向量的数量，即第一数量。进而，可以根据第一数量和全部第一特征向量的总数量

的比值。距离阈值可以根据实际的样本情况进行任意设置。

示例性地，分布数据还可以包括分布方差。即该第一向量组中的第一特征向量的方差。可以理解，方差可以反映每个第一向量组中第一特征向量的离散程度。此步骤可以分别统计每个第一向量组的权重和方差，作为分布数据。

5 步骤 S140，基于参考图像的、第一预设数量的参考向量组各自的分布数据，校正多个第一向量组的分布数据。其中，参考向量组与第一向量组一一对应。

第一预设数量的参考向量组各自的分布数据可以是预先得到的。可以采用任意合适的方法得到参考向量组各自的分布数据，本申请不对其进行限制。

10 参考图像和第一图像可以是同一类型产品的图像。示例性地，以前述 50 个第一图像是对应晶圆新产品的一种缺陷类别（例如缺陷 A）的缺陷图像为例，此步骤中的参考图像可以是以前生产的或者当前正在生产的、与上述晶圆新产品的缺陷特征相似度较高的大批量晶圆产品的缺陷图像。换言之，参考图像和第一图像中缺陷的缺陷类别可以是类似的。例如，参考图像的缺陷类别可以是缺陷 A'。

15 以第一预设数量是 5 为例，可以基于参考图像的 5 个参考向量组各自的分布数据，校正 5 个第一向量组的分布数据。5 个第一向量组的分布数据也可以通过与上述步骤 S110 至 S130 的方法得到。例如，参考图像是从大批量晶圆产品的缺陷图像中选取的一种缺陷类别的图像。大批量晶圆产品的缺陷图像也可以是经人工标注出缺陷类别的参考图像。每种缺陷类别的参考图像可以包括 1000 个。可以利用上述特征提取网络对每种缺陷类别的 1000 个参考图像进行特征提取得到 1000 个参考特征向量。然后，可以采用上述步骤 S120 和步骤 S130 类似的方法针对每种缺陷类别的 1000 个参考图像的参考特征向量聚类出 5 组第二向量组，并得到每组第二向量组各自的分布数据和第二代表特征向量。然后，可以采用各种合适的方法选取对应一种缺陷类别（例如缺陷 A'）的 1000 个参考图像的 5 个第二向量组，作为 1000 个参考图像的 5 个参考向量组。并根据这 5 个参考向量组的分布数据针对步骤 S130 中得到的 50 个第一图像的 5 个第一向量组的分布数据分别进行校准。

25 可以采用任意合适的方法，校正 5 个第一向量组的分布数据。示例性地，可以将 5 个参考向量组与 5 个第一向量组进行匹配，然后利用参考向量组的分布数据对应替换匹配的第一向量组的分布数据。在此步骤中，可以调整全部第一向量组的分布数据，也可以仅调整部分第一向量组的分布数据。

示例性地，对于分布数据包括权重和方差的情况，可以根据参考向量组的权重和方差对匹配的第一向量组的权重和方差分别校正，也可以仅校正其中一者。

步骤 S150，将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，作为多个第一图像的特征表示数据。

可以存储至少部分第一向量组的第一代表特征向量和校正后的分布数据，作为多个第一图像的特征表示数据。以第一预设数目是 5 为例，可以将 5 个第一向量组的第一代表特征向量和校正后的分布数据作为 50 个第一图像的特征表示数据。也可以将 5 个第一向量组中的 3 个或 4 个第一向量组的第一代表特征向量和校正后的分布数据作为 50 个第一图像的特征表示数据。

5 根据上述技术方案，利用参考图像的多个参考向量组各自的分布数据，校正第一图像的第一向量组的分布数据，并根据校正后的分布数据和每个第一向量组的第一代表特征向量作为多个第一图像的特征表示数据。由于校正后的分布数据和每个第一向量组的第一代表特征向量能够准确代表多个第一图像中目标对象的特征。从而，利用这些特征表示数据可以快速且准确地实现对同类型的图像的自动分类。并且，上述方案在第一图像有限的场景下也依然可以适用，方案的适用范围更广泛。

10 示例性地，分布数据包括权重。基于参考图像的、第一预设数量的参考向量组各自的分布数据，校正多个第一向量组的分布数据，包括：从第一预设数量的第一向量组的权重中选取第二预设数量的权重，其中，所选取的权重大于未被选取的权重；利用第一参考向量组的权重，对应地替换所选取的权重。

 如上所述，数值较大的权重，对应出现频率较高的缺陷类别。对于出现频率较高的缺陷而言，调整其权重值具有更好的参考意义。例如第一预设数量是 5，则对多个第一图像可以得到对应 5 个第一向量组的 5 个权重，可以将 5 个向量组按照权重按大小进行排序，权重大的在先，而权重小的在后。同样地，对参考图像的 5 个参考向量组按照的权重的大小进行排序。

 第二预设数量小于或等于第一预设数量。以第一预设数量是 5 为例，第二预设数量可以是小于或等于 5 的任意数目。例如，第二预设数量是 3。在此步骤中，可以从 5 个第一向量组中选取排序在前的 3 个第一向量组，将这 3 个第一向量组的权重作为待校正权重。同样从 5 个参考向量组中选取排序在前的 3 个参考向量组，将这 3 个参考向量组的权重作为目标权重。然后，可以将 3 个目标权重一一替换 3 个待校正权重。例如，5 个参考向量组的权重分别为 0.3、0.2、0.2、0.15、0.15，而 5 个第一向量组的权重分别为 0.5、0.2、0.15、0.1、0.05，则校正后的 5 个第一向量组的权重分别为 0.3、0.2、0.2、0.1、0.05。在未示出的实施例中，也可以根据情况，替换最大的 2 个或 4 个权重。这样可以对分布较为集中的小样本缺陷类别的特征表示数据分别进行校正，使最终的分类结果更加准确。

25 在其他的示例中，分布数据包括分布方差和权重。基于参考图像的、第一预设数量的参考向量组各自的分布数据，校正多个第一向量组的分布数据，还可以包括：根据多个第一向量组各自的权重，将第二预设数量的参考向量组的分布方差对应替换第二预设数量的第一向量组的分布方差。例如，可以将 3 个权重较大的参考向量组的方差对应替换 3 个权重较大的第一向量组的方差。

 示例性地，步骤 S150 将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，

作为多个第一图像的特征表示数据，包括：将被替换权重的第一向量组的权重和第一代表特征向量作为多个第一图像的特征表示数据。例如，可以将排序在先的 3 个第一向量组的权重和第一代表特征向量作为 50 个第一图像的特征表示数据。

示例性地，该方法 100 还包括步骤 S161、步骤 S162、步骤 S163、步骤 S164、步骤 S165、步骤 S166 和步骤 S167。

步骤 S161，计算多个第一特征向量的统计值。以第一图像是晶圆新产品的缺陷图像为例。可以计算当前类别缺陷的 50 个第一图像的第一特征向量的均值向量，表示缺陷的一般特征，如灰度变化。计算当前类别缺陷的 50 个第一图像的第一特征向量的方差向量，表示缺陷的频率特征，如高低频的变化。计算当前类别缺陷的 50 个第一图像的第一特征向量的协方差，表示缺陷的其他属性，比如纹理和色彩的变化。

步骤 S162，利用特征提取网络对多组参考图像中每个参考图像进行特征提取，以得到多个参考特征向量，其中，不同组的参考图像属于不同类型。

如前所述，可以收集大量的缺陷数据，并进行相应的分类标注。例如，缺陷数据可以使用以前生产的大批量产品的缺陷数据，或者当前正在生产的大批量产品的缺陷数据。对于收集到的大量的缺陷数据，可以通过人工对缺陷进行分类标注类别标签。并可以将标注缺陷类别的缺陷样本作为不同类型的参考图像。例如，预先收集并标注了 10 类缺陷样本，每类缺陷样本包含 1000 个图像，则可以将每类缺陷样本的 1000 个图像作为一组参考图像。可以采用图 2 中示出的经训练的两个级联设置的 resnet18 特征提取网络对这 10 组参考图像进行特征提取，得到与 10000 个参考图像一一对应的 10000 个参考特征向量。

如上所述，对于一些差异较为明显的缺陷，例如缺角和污渍，两者可以很容易区分类别。因此，可以将不同类型的参考图像进行分组，对每一组的每个参考图像进行特征提取，从而生成多组参考特征向量。

步骤 S163，针对每组参考图像，对该组参考图像的参考特征向量进行聚类处理，以获取代表该组参考图像的、第一预设数量的第二向量组和代表每个第二向量组的第二代表特征向量。

可以采用与前述步骤 S120 类似的方法，针对每组参考图像中的 1000 个参考特征向量进行聚类处理，得到 5 个第二向量组以及每个第二向量组的第二代表特征向量。

步骤 S164，对于每个第二向量组，计算该第二向量组的分布数据。此步骤与前述步骤 S140 的方法类似，在此不再赘述。

步骤 S165，针对每组参考图像，计算该组参考图像的参考特征向量的统计值。例如，计算每组参考图像的 1000 个参考特征向量的均值向量、方差向量和协方差。

步骤 S166, 根据多个第一特征向量的统计值以及每组参考图像的参考特征向量的统计值, 计算多个第一图像与每组参考图像的距离。步骤 S167, 根据多个第一图像与每组参考图像的距离, 从多组参考图像中选择一组参考图像, 并将代表所选择的该组参考图像的第二向量组确定为参考向量组。

示例性地, 可以通过欧式距离, 根据针对多个第一特征向量的统计值, 如特征均值、方差和协方差与针对每组参考图像的参考特征向量的统计值, 分别计算相关系数, 计算公式如下:

$$Y = a(\text{distance}<\text{均值}>) + b(\text{distance}<\text{方差}>) + c(\text{distance}<\text{协方差}>)$$

A、b、c 可以表示加权系数, 其可以根据实际需求进行任意设置。例如, a、b、c 的取值分别为 0.5、0.3、0.2。示例性地, 可以利用上述公式分别计算多个第一图像所代表的第一缺陷类别与各组参考图像所代表的第二缺陷类别的距离, 将对应最小距离的第二缺陷类别的一组参考图像, 作为校准参考图像。

10 并可以将代表校准参考图像的、第一预设数量的第二向量组, 作为对应的参考向量组。换言之, 可以确定与上述 50 个新晶圆产品的缺陷类别相似的大批量晶圆产品的缺陷类别。

这样, 可以根据统计数据确定多个第一图像和每组参考图像所属类型的对应关系。对于存在多组第一图像的情况, 则可以确定每组第一图像匹配 (类别相似) 的一组参考图像。即可以通过将第二向量组与第一向量组匹配, 从而确定与第一向量组类别相似的参考向量组。进而可以利用参考向量组的分布数据对第一向量组的分布数据进行校准。

15 根据本申请另一方面, 还提供一种图像样本分类方法。图 3 示出根据本申请一个实施例的图像样本分类方法 300 的示意性流程图。该分类方法 300 包括:

步骤 S310, 利用特征提取网络对图像样本进行特征提取, 以得到样本特征向量。示例性地, 图像样本可以是新产品图像的待进行缺陷分类的缺陷图像数据。可以采用上述经训练的两个级联设置的
20 resnet18 特征提取网络提取图像样本的特征, 得到样本特征向量。

步骤 S320, 基于样本特征向量和多组第一图像中每组第一图像的特征表示数据, 确定图像样本所属类型。其中, 每组第一图像的特征表示数据是利用上述图像处理方法获取的。不同组的第一图像属于不同类型。可以将样本特征向量与每组第一图像的特征表示数据进行匹配, 根据匹配结果确定图像样本所属类型。

25 在工业生产领域, 广泛应用机器学习进行产品定位和缺陷检测等操作。机器学习算法可分为监督学习和无监督学习, 监督学习例如对产品的某一种缺陷, 例如污渍等进行检测, 通过将推理结果与人工标定的结果进行比较, 从而使按照用户的需求输出结果。无监督学习则可以仅由用户确定分类数量, 机器可以对缺陷进行自行判断, 将种类接近的缺陷归入一组, 从而确定多组不同种类的缺陷, 不同的组中缺陷差异很大, 而同一组中的缺陷差异较小。这样在检测缺陷时, 机器可以自行对缺陷进行分组, 并根据

向每个分组中新增加的样本，调整对应分组的中心，从而使缺陷的分组更加精准。换言之，随着样本量的增加，样本的分布更加接近正态分布，数据更不易受到异常样本数据的影响，可靠性更高。但对于样本量很少的样本，沿用上述无监督学习的方法，可能因为个别样本的值偏离均值较大，从而严重影响对缺陷的分类准确程度。而根据本申请实施例的上述图像样本分类方法，通过对积累有大量数据产品的特征提取和训练，获取其每种缺陷类别的分布，在新的类似或相关产品中，使用获取的特征分布去进行校准，有效解决现有技术中针对小样本分类效果不佳的问题。并且划分的类别更加准确。

示例性地，基于样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，包括：根据样本特征向量与每组第一图像的特征表示数据，计算每组第一图像与图像样本之间的第一距离，并得到多个第一距离。将多个第一距离中的最小值所对应的一组第一图像所属的类型确定为图像样本所属类型。

如上所述，示例性地，一组第一图像对应的 5 个第一向量组各自的第一代表特征向量分别为 A、B、C、D、E，对应的校正后的权重分别为 0.3、0.2、0.2、0.1、0.05，特征表示数据可以包括部分第一向量组的权重和这些第一向量组的第一代表特征向量。例如，一组第一图像的特征表示数据包括第一代表特征向量 A、B、C 和对应的权重 0.3，0.2，0.2。此步骤中可以采用加权平均的方法计算每组第一图像与图像样本之间的第一距离。图像样本的样本特征向量例如是 x，则图像样本与该组第一图像之间的第一距离可以等于 $0.3 * (x-A) + 0.2 * (x-B) + 0.2 * (x-C)$ 。并且，可以采用类似的方法计算得到每组第一图像与图像样本之间的第一距离。并可以将最小的第一距离所对应的一组第一图像所属的缺陷类别作为当前图像样本所属的缺陷类别。通过计算并比较第一距离，可以快速判断图像样本所属类型。这种方案执行逻辑简单，计算量更小，从而这种分类方法分类效率更高，分类结果也更精准。

图 4 示出根据本申请另一实施例的图像样本分类方法的流程图。如图所示。在第一阶段，可以收集并分析老产品的缺陷数据。首先，可以收集大量的缺陷数据，并进行相应的分类标注。可以使用以前生产的大批量产品的缺陷数据，或者当前正在生产的大批量产品的缺陷数据，可以收集到大量的缺陷数据集 A，通过人工对缺陷进行分类标注类别标签（缺陷类别可以被称作旧缺陷）。然后，可以训练特征提取和分类网络。特征提取和分类网络由 3 个子网络分组成，包括两个 resnet18 特征提取网络和一个分类网络。resnet18_1 网络可以使用 Imagenet 数据预训练的权重。使用 S1 中人工分类标注后的缺陷数据集 A 训练特征提取和分类网络。训练完成后，得到最优的特征提取和分类网络。然后，可以对缺陷数据集 A 中的每个类别的缺陷图像的特征进行统计分布。缺陷数据集 A 中例如包括对应 3 种缺陷类别的 3 组参考图像。具体地，针对每一种缺陷类别的多个参考图像，通过训练好的两个 resnet18 特征提取网络进行特征提取，得到多个参考特征向量。统计多个参考特征向量的均值向量、方差向量和协方差。然后，将多

个参考特征向量进行聚类处理，得到 5 个第二向量组和每个第二向量组的第二代表特征向量、权重和方差。之后，可以将所有旧缺陷类别和每个旧缺陷类别中的 5 个参考向量组的上述统计信息，存入当前产品 recipe 当中。

在第二阶段，可以收集并分析新产品的缺陷数据。例如，针对新产品的每种缺陷数据可以收集到 30-50 张，就可以满足本方案的小样本分类对于缺陷数据的需求。新产品缺陷数据集 B 例如由人工标注出 3 种缺陷类别（缺陷类别可以被称作新缺陷）。按照上述方法利用上述训练好的特征提取网络对每种新缺陷的多个第一图像进行特征提取，得到多个第一特征向量。统计多个第一特征向量的均值向量、方差向量和协方差。然后，对多个第一特征向量进行聚类处理，得到 5 个第一向量组。统计每个第一向量组的第一代表特征向量、权重和方差。由于新产品的样本数量少，所以得到的特征分布可以被称作稀疏特征分布。

在第三阶段，可以利用老产品的特征分布对新产品的特征分布进行校准。首先，可以计算每类新缺陷和每类旧缺陷的距离。具体地，可以根据每类新缺陷的多个第一特征向量的均值向量、方差向量和协方差与每类老缺陷的多个参考特征向量的均值向量、方差向量和协方差，计算每类新缺陷与每类老缺陷之间的距离。将与每类新缺陷距离最小的老缺陷确定为相似缺陷，将相似缺陷对应的 5 个第二向量组确定为该类新缺陷图像的参考向量组。对于每类新缺陷，利用对应的 5 个参考向量组的权重或方差校准该新缺陷的 5 个第一向量组的权重或方差。例如，可以直接利用权重值最大的 3 个参考向量组的权重对应替换权重值最大的 3 个第一向量组的权重。由此，得到 3 类新缺陷各自的特征表示数据。

在第四阶段，推理待分类的缺陷图像，划分类别。当新产品检测到一个缺陷，送入一张缺陷图像数据，使用特征提取网络进行特征提取后得到样本特征向量。然后，根据 3 类新缺陷各自的特征表示数据和样本特征向量，计算当前缺陷图像与每个新缺陷的距离，将距离最小的新缺陷类别，判定为当前缺陷图像的缺陷类别。

根据本申请另一方面，还提供一种图像处理装置。图 5 示出根据本申请实施例的图像处理装置 500 的示意性框图。如图所示，图像处理装置 500 包括：

第一特征提取模块 510，用于利用特征提取网络对多个第一图像分别进行特征提取，以得到多个第一特征向量；

聚类模块 520，用于对多个第一特征向量进行聚类处理，以获取代表多个第一图像的、第一预设数量的第一向量组和代表每个第一向量组的第一代表特征向量；

第一计算模块 530，用于对于每个第一向量组，计算该第一向量组的分布数据，其中分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差；

校正模块 540, 用于基于参考图像的、第一预设数量的参考向量组各自的分布数据, 校正多个第一向量组的分布数据, 其中, 参考向量组与第一向量组一一对应;

第二计算模块 550, 用于将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据, 作为多个第一图像的特征表示数据。

5 根据本申请另一方面, 还提供一种图像样本分类装置。图 6 示出根据本申请实施例的图像样本分类装置 600 的示意性框图。如图所示, 图像样本分类装置 600 包括:

第二特征提取模块 610, 用于利用特征提取网络对图像样本进行特征提取, 以得到样本特征向量;

分类模块 620, 用于基于样本特征向量和多组第一图像中每组第一图像的特征表示数据, 确定图像样本所属类型, 其中, 每组第一图像的特征表示数据是利用上述图像处理方法获取的, 不同组的第一图
10 像属于不同类型。

根据本申请的另一方面, 还提供一种电子设备。图 7 示出根据本申请实施例的电子设备 700 的示意性框图。如图所示, 电子设备 700 包括处理器 710 和存储器 720。存储器 720 中存储有计算机程序指令, 计算机程序指令被处理器 710 运行时用于执行上述图像处理方法 100。

根据本申请的另一方面, 还提供一种存储介质。在存储介质上存储了程序指令, 程序指令在运行时
15 用于执行上述图像处理方法 100。所述存储介质例如可以包括可擦除可编程只读存储器(EPROM)、便携式只读存储器(CD-ROM)、USB 存储器、或者上述存储介质的任意组合。所述存储介质可以是一个或多个计算机可读存储介质的任意组合。

本领域普通技术人员通过阅读上述有关图像处理方法 100 和图像样本分类方法 300 的相关描述, 可以理解上述图像处理装置 500、图像样本分类装置 600、电子设备 700 和存储介质的具体实施方案和有益效果, 为了简洁, 在此不再赘述。
20

实施例:

实施例 1: 一种图像处理方法, 其中, 包括:

利用特征提取网络对多个第一图像分别进行特征提取, 以得到多个第一特征向量;

对所述多个第一特征向量进行聚类处理, 以获取代表所述多个第一图像的、第一预设数量的第一向
25 量组和代表每个第一向量组的第一代表特征向量;

对于每个第一向量组, 计算该第一向量组的分布数据, 其中所述分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差;

基于参考图像的、所述第一预设数量的参考向量组各自的分布数据, 校正所述多个第一向量组的分布数据, 其中, 所述参考向量组与所述第一向量组一一对应; 以及

将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，作为所述多个第一图像的特征表示数据。

实施例 2：根据实施例 1 所述的方法，其中，所述特征提取网络包括级联的两个特征提取网络。

实施例 3：根据实施例 1 或 2 所述的方法，其中，所述分布数据为权重，所述计算该第一向量组的

5 分布数据，包括：

计算该第一向量组中的所有第一特征向量与该第一向量组中的第一代表特征向量之间的距离；

计算第一数量与所述多个第一特征向量的总数量之间的比值，以作为该第一向量组的权重；其中，所述第一数量是与该第一向量组中的第一代表特征向量之间的距离小于距离阈值的第一特征向量的数量。

10 实施例 4：根据实施例 1-3 中任一实施例介绍的方法，其中，所述分布数据包括权重，所述基于参考图像的、所述第一预设数量的参考向量组各自的分布数据，校正所述多个第一向量组的分布数据，包括：

从所述第一预设数量的第一向量组的权重中选取第二预设数量的权重，其中所选取的权重大于未被选取的权重；

15 利用第一参考向量组的权重，对应地替换所选取的权重。

实施例 5：根据实施例 1-4 中任一实施例介绍的方法，其中，所述处理方法还包括：

计算所述多个第一特征向量的统计值；

利用特征提取网络对多组参考图像中每个参考图像进行特征提取，以得到多个参考特征向量，其中，不同组的参考图像属于不同类型；

20 针对每组参考图像，对该组参考图像的参考特征向量进行聚类处理，以获取代表该组参考图像的、第一预设数量的第二向量组和代表每个第二向量组的第二代表特征向量；

对于每个第二向量组，计算该第二向量组的分布数据；

针对每组参考图像，计算该组参考图像的参考特征向量的统计值；

根据所述多个第一特征向量的统计值以及每组参考图像的参考特征向量的统计值，计算多个第一图

25 像与每组参考图像的距离；

根据所述多个第一图像与每组参考图像的距离，从所述多组参考图像中选择一组参考图像，并将代表所选择的该组参考图像的第二向量组确定为所述参考向量组。

实施例 6：一种图像样本分类方法，其中，包括：

利用特征提取网络对图像样本进行特征提取，以得到样本特征向量；

基于所述样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，其中，每组第一图像的特征表示数据是利用实施例 1-5 中任一实施例所述的图像处理方法获取的，不同组的第一图像属于不同类型。

实施例 7：根据实施例 6 中介绍的方法，其中，所述基于所述样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，包括：

根据所述样本特征向量与每组第一图像的特征表示数据，计算每组第一图像与所述图像样本之间的第一距离，并得到多个第一距离；

将所述多个第一距离中的最小值所对应的一组第一图像所属的类型确定为所述图像样本所属类型。

实施例 8：根据实施例 6 或 7 中介绍的方法，其中，所述第一图像是产品的图像，第一图像所属类型根据产品的缺陷的类型来确定。

实施例 9：一种图像处理装置，其中，包括：

第一特征提取模块，用于利用特征提取网络对多个第一图像分别进行特征提取，以得到多个第一特征向量；

聚类模块，用于对所述多个第一特征向量进行聚类处理，以获取代表所述多个第一图像的、第一预设数量的第一向量组和代表每个第一向量组的第一代表特征向量；

第一计算模块，用于对于每个第一向量组，计算该第一向量组的分布数据，其中所述分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差；

校正模块，用于基于参考图像的、所述第一预设数量的参考向量组各自的分布数据，校正所述多个第一向量组的分布数据，其中，所述参考向量组与所述第一向量组一一对应；

第二计算模块，用于将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，作为所述多个第一图像的特征表示数据。

实施例 10：一种图像样本分类装置，其中，包括：

第二特征提取模块，用于利用特征提取网络对图像样本进行特征提取，以得到样本特征向量；

分类模块，用于基于所述样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，其中，每组第一图像的特征表示数据是利用实施例 1-5 中任一实施例所述的图像处理方法获取的，不同组的第一图像属于不同类型。

实施例 11：一种电子设备，包括处理器和存储器，其中，所述存储器中存储有计算机程序指令，所述计算机程序指令被所述处理器运行时用于执行如实施例 1-5 中任一实施例所述的图像处理方法和/或如实施例 6-8 任一实施例所述的图像样本分类方法。

实施例 12：一种存储介质，在所述存储介质上存储了程序指令，其中，所述程序指令在运行时用于执行如实施例 1-5 任一实施例所述的图像处理方法和/或如实施例 6-8 任一实施例所述的图像样本分类方法。

5 尽管这里已经参考附图描述了示例实施例，应理解上述示例实施例仅仅是示例性的，并且不意图将本申请的范围限制于此。本领域普通技术人员可以在其中进行各种改变和修改，而不偏离本申请的范围和精神。所有这些改变和修改意在包括在所附权利要求所要求的本申请的范围之内。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现
10 现所描述的功能，但是这种实现不应认为超出本申请的范围。

在本申请所提供的几个实施例中，应该理解到，所揭露的设备和方法，可以通过其它的方式实现。例如，以上所描述的设备实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个设备，或一些特征可以忽略，或不执行。

15 在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本申请的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

类似地，应当理解，为了精简本申请并帮助理解各个发明方面中的一个或多个，在对本申请的示例性实施例的描述中，本申请的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，
20 并不应将该本申请的方法解释成反映如下意图：即所要求保护的本申请要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如相应的权利要求书所反映的那样，其发明点在于可以用少于某个公开的单个实施例的所有特征的特征来解决相应的技术问题。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本申请的单独实施例。

本领域的技术人员可以理解，除了特征之间相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由
25 提供相同、等同或相似目的的替代特征来代替。

此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本申请的范围之内并且形成不同的实施例。

例如，在权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

本申请的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本申请实施例的图像处理装置 500、图像样本分类装置 600 中的一些模块的一些或者全部功能。本申请还可以实现为用于执行这里所描述的方法的一部分或者全部的装置程序（例如，计算机程序和计算机程序产品）。这样的实现本申请的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

10 15 需要注意的是上述实施例对本申请进行说明而不是对本申请进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本申请可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

以上所述，仅为本申请的具体实施方式或对具体实施方式的说明，本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。本申请的保护范围应以权利要求的保护范围为准。

权利要求

1.一种图像处理方法，其特征在于，包括：

利用特征提取网络对多个第一图像分别进行特征提取，以得到多个第一特征向量；

对所述多个第一特征向量进行聚类处理，以获取代表所述多个第一图像的、第一预设数量的第一向

5 量组和代表每个第一向量组的第一代表特征向量；

对于每个第一向量组，计算该第一向量组的分布数据，其中所述分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差；

基于参考图像的、所述第一预设数量的参考向量组各自的分布数据，校正所述多个第一向量组的分布数据，其中，所述参考向量组与所述第一向量组一一对应；以及

10 将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，作为所述多个第一图像的特征表示数据。

2.如权利要求 1 所述的图像处理方法，其特征在于，所述特征提取网络包括级联的两个特征提取网络。

3.如权利要求 1 所述的图像处理方法，其特征在于，所述分布数据为权重，所述计算该第一向量组的分布数据，包括：

计算该第一向量组中的所有第一特征向量与该第一向量组中的第一代表特征向量之间的距离；

计算第一数量与所述多个第一特征向量的总数量之间的比值，以作为该第一向量组的权重；其中，所述第一数量是与该第一向量组中的第一代表特征向量之间的距离小于距离阈值的第一特征向量的数量。

20 4.如权利要求 1 所述的图像处理方法，其特征在于，所述分布数据包括权重，所述基于参考图像的、所述第一预设数量的参考向量组各自的分布数据，校正所述多个第一向量组的分布数据，包括：

从所述第一预设数量的第一向量组的权重中选取第二预设数量的权重，其中所选取的权重大于未被选取的权重；

利用第一参考向量组的权重，对应地替换所选取的权重。

25 5.如权利要求 1 至 4 任一项所述的图像处理方法，其特征在于，所述处理方法还包括：

计算所述多个第一特征向量的统计值；

利用特征提取网络对多组参考图像中每个参考图像进行特征提取，以得到多个参考特征向量，其中，不同组的参考图像属于不同类型；

针对每组参考图像，对该组参考图像的参考特征向量进行聚类处理，以获取代表该组参考图像的、

第一预设数量的第二向量组和代表每个第二向量组的第二代表特征向量；

对于每个第二向量组，计算该第二向量组的分布数据；

针对每组参考图像，计算该组参考图像的参考特征向量的统计值；

根据所述多个第一特征向量的统计值以及每组参考图像的参考特征向量的统计值，计算多个第一图

5 像与每组参考图像的距离；

根据所述多个第一图像与每组参考图像的距离，从所述多组参考图像中选择一组参考图像，并将代表所选择的该组参考图像的第二向量组确定为所述参考向量组。

6.一种图像样本分类方法，其特征在于，包括：

利用特征提取网络对图像样本进行特征提取，以得到样本特征向量；

10 基于所述样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，其中，每组第一图像的特征表示数据是利用权利要求 1 至 5 任一项所述的图像处理方法获取的，不同组的第一图像属于不同类型。

7.如权利要求 6 所述的图像样本分类方法，其特征在于，所述基于所述样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，包括：

15 根据所述样本特征向量与每组第一图像的特征表示数据，计算每组第一图像与所述图像样本之间的第一距离，并得到多个第一距离；

将所述多个第一距离中的最小值所对应的一组第一图像所属的类型确定为所述图像样本所属类型。

8.如权利要求 6 或 7 所述的图像样本分类方法，其特征在于，所述第一图像是产品的图像，第一图像所属类型根据产品的缺陷的类型来确定。

20 9.一种图像处理装置，其特征在于，包括

第一特征提取模块，用于利用特征提取网络对多个第一图像分别进行特征提取，以得到多个第一特征向量；

聚类模块，用于对所述多个第一特征向量进行聚类处理，以获取代表所述多个第一图像的、第一预设数量的第一向量组和代表每个第一向量组的第一代表特征向量；

25 第一计算模块，用于对于每个第一向量组，计算该第一向量组的分布数据，其中所述分布数据包括该第一向量组的权重和/或该第一向量组中的第一特征向量的分布方差；

校正模块，用于基于参考图像的、所述第一预设数量的参考向量组各自的分布数据，校正所述多个第一向量组的分布数据，其中，所述参考向量组与所述第一向量组一一对应；

第二计算模块，用于将校正后的至少部分第一向量组的第一代表特征向量和所校正的分布数据，作

为所述多个第一图像的特征表示数据。

10.一种图像样本分类装置，其特征在于，包括：

第二特征提取模块，用于利用特征提取网络对图像样本进行特征提取，以得到样本特征向量；

5 分类模块，用于基于所述样本特征向量和多组第一图像中每组第一图像的特征表示数据，确定图像样本所属类型，其中，每组第一图像的特征表示数据是利用权利要求 1 至 5 任一项所述的图像处理方法获取的，不同组的第一图像属于不同类型。

11.一种电子设备，包括处理器和存储器，其特征在于，所述存储器中存储有计算机程序指令，所述计算机程序指令被所述处理器运行时用于执行如权利要求 1 至 5 任一项所述的图像处理方法和/或如权利要求 6 至 8 任一项所述的图像样本分类方法。

10 12.一种存储介质，在所述存储介质上存储了程序指令，其特征在于，所述程序指令在运行时用于执行如权利要求 1 至 5 任一项所述的图像处理方法和/或如权利要求 6 至 8 任一项所述的图像样本分类方法。